小山工業高等専門学校 電気電子創造工学科 平成 30 年度 卒業論文

広域センサネットワークを用いた トラッキングサービスにおける サービスに不要なプライバシデータ流出比の定量評価

Quantitative Evaluation of Ability for Privacy Protection Using Simulator of Local Production for Local Consumption Type Networks in Wide Area Sensor Tracking Service

> 電気電子創造工学科 田村 崚 Department of Innovative Electrical Electronic Engineering Ryo TAMURA

> > 指導教員 干川 尚人 Naoto HOSHIKAWA

指導教員

目次

- 第1章 序章
- 第2章 既存研究
- 2.1.クラウドモデル
- 2.2.LPLC モデル
- 第3章 数理モデル提案
- 3.1.クラウドモデルの数理モデル化
- 3.2.LPLC モデルの数理モデル化
- 第4章 シミュレーション実験
- 4.1.提案手法
- 4.1.1.プロトタイプシミュレータについて
- 4.1.2. オープンデータについて
- 4.1.3.シミュレーション方式
- 4.2.シミュレータの設計
- 4.2.1.サービスモデル
- 4.2.2. 対象者設定
- 4.3.シミュレーション内容
- 4.4.シミュレーション結果
- 4.5.考察
- 第5章 まとめ

第1章 序章

近年、センサなどを備えた電子機器をインターネット接続可能にする技術、いわゆる Internet of Things (IoT)技術の発展がめざましい。また公共の場に設置される IoT 機器はその維持費などの観点から、サービス固有のものではなく公共の所有物となると言われている (IoT 機器のオープン化)。IoT 技術の発展及びオープン化、また画像処理技術の向上によって、広域センサネットワークを用いた情報処理サービス (例えば監視カメラを用いた地域パトロールのような見守りサービス) がより広域かつ高精度になると期待される。

一方で、このような広域でのセンシングが必要なサービスは多くのプライバシ情報を収集してしまう問題がある。事実、顔などの生体情報の提供に抵抗のある利用者は特に多く[1]、もし有用なサービスが提案されたとしても社会実装に理解の得られない可能性がある。この課題に対して、サービスモデルに共通のプライバシ保護能力についての指標を設定し定量評価をする、またプライバシ保護に着目したサービスモデルを提案するといった解決アプローチがある。

プライバシ保護に着目したサービスモデルとして、本研究チームは地産地消型ネットワークアーキテクチャ(Local Production for Local Consumption Network Architecture,以下 LPLCモデルと記載)を提案している[2].

サービスモデルに共通のプライバシ保護能力についての指標について、既存研究には生まれる情報量から個人特定に至りやすいか否かでプライバシ量の評価を図るもの[3]があるが、広域センサネットワークを用いた見守りサービスに適したモデルはない。本稿では多くの人にサービスの安全性を伝えられる指標として、「不要なプライバシデータ流出比」を提案し、その数理モデルを示す。また既存サービスモデルと LPLC モデルの不要なプライバシデータ流出比を算出するシミュレータを作成し、この二つを比較検討する。

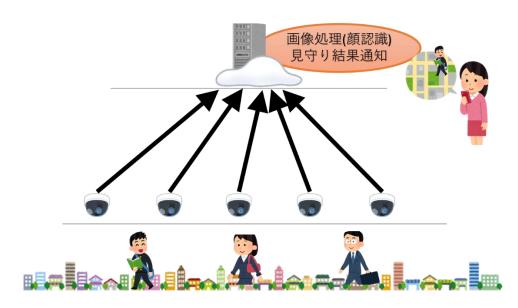
第2章 既存研究

広域センサネットワークを用いた見守りサービスを実現する手法はいくつか提案されている.この章では見守りサービスを実現する既存研究のうち、もっともメジャーであるクラウドモデルと本研究チームの提案したLPLCモデルの2つを紹介する.

2.1. クラウドモデル

既存の見守りサービスは主にクラウドモデルを使って実現されている[3][4]. クラウドモデルではあらゆるセンサデータをクラウド上にあるサーバに送信し、情報処理はそのサーバ上で行っている. その処理結果によって見守り依頼者へ情報を送信する.

クラウドモデルはあらゆるデータを失うことなく 1 か所で処理できるため汎用性に優れている.



2.2. LPLC モデル

LPLC モデルの特徴として、動的に見守り対象者がセンシングされているセンサを選択する事が挙げられる。こうすることでサービスの実現には不要な人物のセンシングを省くことが可能になり、クラウドモデルよりもプライバシ保護が可能になる。

このモデルの実現にあたり、無線アクセスポイントの持つ物理的な接続限界距離に着目する。センサの近くに無線アクセスポイントを設置し、センサとローカルネットワークを形成させる。無線アクセスポイントの持つ物理的な接続可能範囲を、以降アクセス可能範囲と記載する。見守り対象者には無線アクセスポイント接続機能を持つ計算ノード(以下 LPLC ノードと記載)を携帯させる(例としてスマートフォンが挙げられる)。 LPLC ノードが無線アクセスポイントへ接続した時、同ローカルネットワークにあるセンサデータを LPLC ノードへ送信し、LPLC ノードはそのデータを処理する。その処理結果によって見守り依頼者へ情報を送信する。

LPLC モデルは必要なデータのみを収集するため、あらゆるデータを収集するクラウドモデルと比べて、不要なプライバシデータの流通を制御できる.





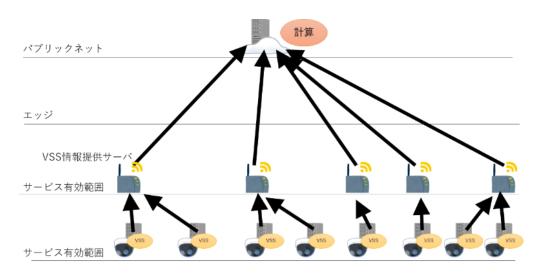
第3章 数理モデル提案

序章で触れた通り、多くの人にサービスの安全性を伝えるには共通の指標を持ち定量評価することが大切である.この章では2章で挙げた既存モデルに対して「不要なプライバシデータ流出比」という値を定義する.

なお、ここでは計算ノードにデータが流入することを流出と定義している.

3.1. クラウドモデルの数理モデル化

クラウドモデルの動作モデルを以下のように考えた.



図クラウドモデルの動作モデル図

サービス領域を内包するように長方形(以下マップと記載)を置き、それを外包するように正方形のセルを敷き詰める。それぞれのセルが持つ対象者の有無、センサ有効範囲内といった情報を各行列の要素に対応させモデル化する。まずn番目のカメラ集合を定義する。この集合は、各セルがカメラのセンサ有効範囲内か否かを格納した行列 $C_n(t)$ と、マップ上でのカメラの撮影方向up,down,right,leftの逆 $C_{d_n}(t)$ を要素に持つ。次にn番目のルータの集合 R_n を定義する。これは属するカメラ集合の集合 RC_n とアクセス可能範囲行列 A_n を要素に持つ。ここでルータ位置とアクセス可能距離をそれぞれ(x,y),rとすると、 A_n は次のように定義できる。

$$[A_n]_{ij} = \begin{cases} 1 & \text{if } x - r \le i \le x + r \text{ and} \\ y - r \le j \le y + r \\ 0 & \text{otherwise} \end{cases}$$
 (1)

最後にn人目の対象者の集合を定義する.この集合は各セルでの対象者の有無を格納した行列 $P_n(t)$ と対象者の向いている方向を表現する行列 $P_{up_n}(t)$ 、 $P_{down_n}(t)$ 、 $P_{right_n}(t)$ 、 $P_{left_n}(t)$

を要素に持つ. それぞれを(2)(3)(4)(5)に示す. 例として, $P_{up_n}(t)_{ij}$ が1であるとは, n人目の対象者が時刻 t の時にセル ij でマップ上方を向いていたことを示している. それぞれの行列の定義を下に示す. また, $P_n(t)$ の総和をP(t)と定義する. 方向を示す行列も同様である. 非対象者集合 $N_n(t)$ も同様に定義する.

$$[P_{up_n}(t)]_{ij} = -\begin{cases} 1 & \text{if } P_n(t)_{ij} \cdot P_n(t-1)_{(i+1)j} \neq 0 \\ 0 & \text{otherwise} \end{cases}$$
 (2)

$$\left[P_{down_n}(t)\right]_{ij} = -\begin{cases} 1 & \text{if } P_n(t)_{ij} \cdot P_n(t-1)_{(i-1)j} \neq 0\\ 0 & \text{otherwise} \end{cases}$$

$$\left[P_{right_n}(t)\right]_{ij} = - \begin{cases} 1 & \text{if } P_n(t)_{ij} \cdot P_n(t-1)_{i(j-1)} \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\left[P_{left_n}(t)\right]_{ij} = \begin{cases} 1 & \text{if } P_n(t)_{ij} \cdot P_n(t-1)_{i(j+1)} \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

サービス開始時間を t_0 ,終了時間を t_n とすると,クラウドモデルにおける NPD 流出比は次式のようになる.

$$\int_{t_0}^{t_n} \sum_{n \in \mathbb{R}} \frac{S_n(N(t))}{S_n(P(t)) + S_n(N(t))} dt$$

なお、 $S_n(P(t))$ は n 番目のルータが時刻 t に対象者全体のセンシング量を返す関数、

$$S_n(P(t)) = \sum_{n \in RC_n} P_{C_{d_n}(t)}(t) \cdot C_n(t)$$

3.2. LPLC モデルの数理モデル化

LPLC の動作モデル図を図1に示す.

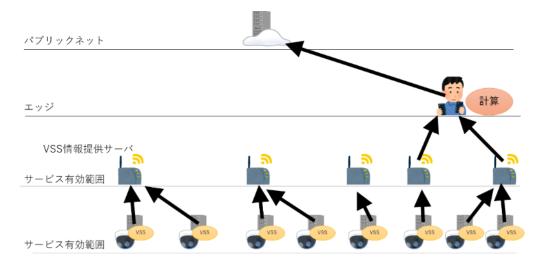


図 1.LPLC の動作モデル図

サービス開始時間を t_0 ,終了時間を t_n とすると,LPLC モデルにおける NPD 流出比は次式のようになる. $\delta(P(t),A_n)$ はn番目のルータのアクセス可能範囲内に時刻tの対象者が存在するか否かを返す関数である.

$$\int_{t_0}^{t_n} \sum_{n \in R} \left\{ \delta(P(t), A_n) \frac{S_n(N(t))}{S_n(P(t)) + S_n(N(t))} \right\} dt$$

$$\delta(P(t), An) = \begin{cases} 1 & \text{If } Tr(P(t) \cdot {}^{t}A_{n}) \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

第4章 シミュレーション実験

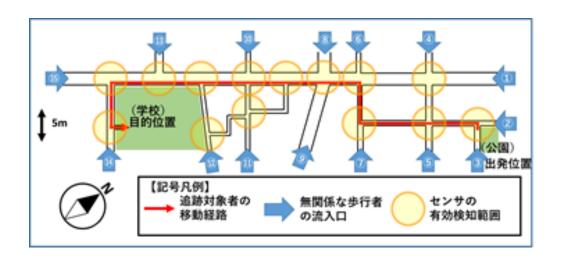
この章では第3章で提案した数理モデルをシミュレータ上で再現し、NPD 流出比を算出する.シミュレータ作成にあたり、見守りサービスの定義や公開されているデータの利用手法についても述べる.

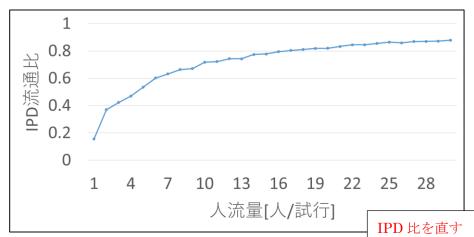
4.1. 提案手法

この節では公開されているデータについて紹介し、どのようにしてシミュレーションへ 利用可能な形式へ変換したかを示す.また、プロトタイプシミュレータについても述べる.

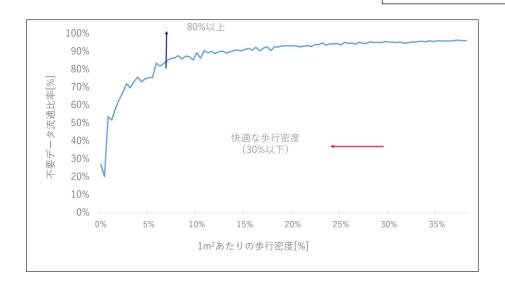
4.1.1.プロトタイプシミュレータについて

以前簡易的なクラウドサービスのシミュレータ[5]を作成し、クラウドモデルの NPD 比について検討を行った.このシミュレーションでは実在する小学校の周囲をモデルにサービス領域を作成(図)し、サービス内容は「公園から学校までの見守り」がクラウドサービスで提供されるとした.見守り対象者(以下対象者と記載)を児童とし、そうでない人物を見守り非対象者(以下非対象者と記載)とした.カメラは全ての交差点に設置し、このシミュレーションではカメラの方向を考えない(数式).人の歩行速度を 1.25[m/sec]、カメラの動作周期を t=1[s]、セル1 辺を 1.25[m]とした.また、追跡対象者は赤線に示す固定のルートを、非対象者は青矢印に示す 15 カ所の流入口から 1 つ選びそこからのルートを立ち止まることなく進む.非対象者が増減した際の NPD 流出比を図に示す.また、横軸を 1m²あたりの歩行密度に直したものを図に示す.この結果から、快適と言われる 30%以下の歩行密度であっても NPD 流出比は 80%を超えていることがわかった.





できればデータをゲットしたい

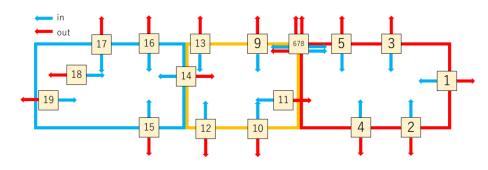


4.1.2.オープンデータについて

プロトタイプでは見守り対象者の動きや人数について自由に設定したため、現実的な NPD 流出比であるかは疑問が残った。そこで本稿では G 空間情報センターの公表している 松江駅構内人流センサデータ (西日本旅客鉄道株式会社) [6] (以下オープンデータと記載) からより現実に即した見守り対象者の動きや人数を取得し、より現実に即した NPD 流出比を得る手法を考案した。オープンデータはセンサ 19 個ごとに「センサ ID、日付、時刻、In、Out、In 累計、Out 累計」が定義されている。また、センシング結果は 1 分毎の合計が記録されている。センサ位置、in/out データは公開されている次の画像(図)に従う。赤い枠で囲われた部分を「シャミネ東」、オレンジ色の枠で囲われた部分を「コンコース」、青い枠で囲われた部分を「シャミネ東」、オレンジ色の枠で囲われた部分を「コンコース」、青い枠で囲われた部分を「シャミネ西」と呼称する。以降簡単のために枠を記載するが、シミュレータの動作には影響しない。なお、本稿では 2018 年 6 月、平日の 5 時から 24 時のデータを扱う。人流センサ 15 番は動作を確認できなかったため扱っていない。



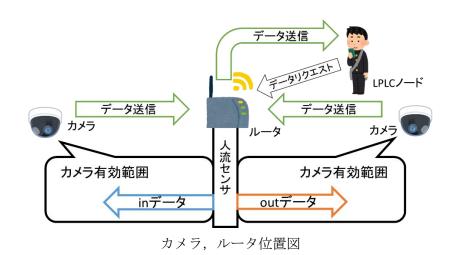
センサ設置位置図



センサ in/out 定義図

4.1.3.シミュレーション方式

オープンデータはあくまで人流センサデータであるため、このままではプライバシデータとして扱えない。そこで in/out 人流のそれぞれを撮影する仮想のカメラを設定し、人流センサデータを画像データへ変換する機構を作成した(図)。カメラは RD-CA213 AHD カメラ220 万画素バレット型赤外線搭載屋外カメラ[7]を想定し、同カメラを販売している会社の公表しているデータ[8]からセンシング可能距離を15m とした。また LPLC モデルにはカメラの近くにルータを設置する必要があるため、オープンデータにおける人流センサと同じ位置にルータを設定した。



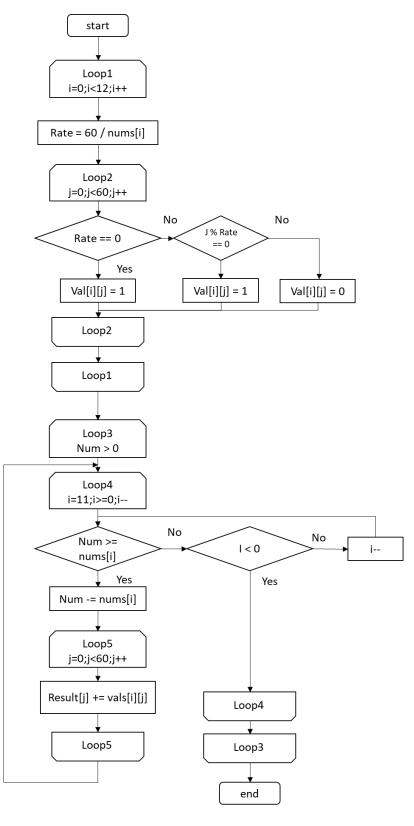
本稿では見守りサービスに用いるカメラのセンシング周期を1秒とする. オープンデータは1分毎に更新されるので、次に示すアルゴリズムを用いてオープンデータを1秒毎のデータへ変換した. アルゴリズムの様子を図に示す.

· ·	
60 の約数	nums[12]={1,2,3,4,5,6,10,12,15,20,30,60}
約数毎の 60 分割	vals[12][60]
入力データ	input
出力データ	result[60]

表1. アルゴリズムに用いる変数の定義

vals[i]には nums[i]を 60 分割したデータが入っている。例えば 20 の vals には $\{1,0,0,1,0,0,1,\dots,1,0,0\}$ と 1,0,0 が 20 回繰り返されるデータを格納している。このように vals には 60 個の要素の合計が nums となりつつ,数字が偏らないように格納されている。input が 0 になるまで次の処理を繰り返すことで vals に v

- ① input より小さい中で、最も大きい nums を選ぶ. 選ばれた nums を仮に nums[i]とする.
- ② input から nums[i]を引き, result に vals[i]を足し合わせる.

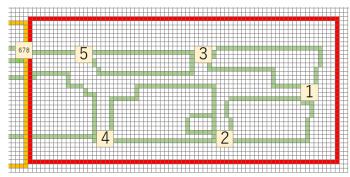


4.2. シミュレータの設計

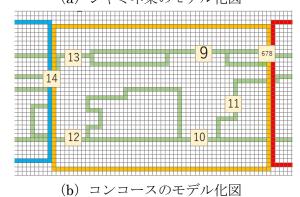
この節ではクラウドモデル及び LPLC モデルを適用する見守りサービスを定義する.

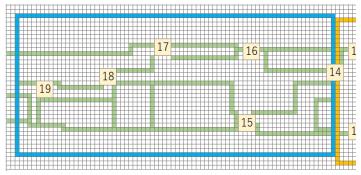
4.2.1.サービスモデル

ここでは広域見守りサービスの例として、松江駅登校者見守りサービスを設定する. サービス領域はオープンデータで示された領域(図)の内、壁に囲まれていない部分とし、サービス内容を「駅を利用する高校生の登校見守り」とした. サービス時間は曜日や学校に依存しにくい登校を想定し、7時から8時とした. 対象者は松江駅を利用して登校する高校生とし、そうでない人物を非対象者とした. オープンデータ全てを非対象者のデータとし、対象者のデータについては次の項で述べる. カメラの位置、方向は図で示したとおりである. 人の歩行速度を1.25[m/sec]、カメラの動作周期をt=1[s]、セル1辺を1.25[m]とした. モデル化された JR 松江駅を図に示す. サービス領域外を灰色で、センシング可能範囲を緑色で可視化したものを図に示す.

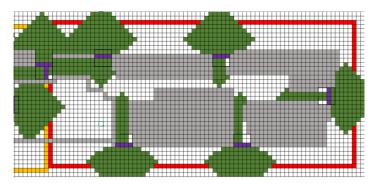


(a) シャミネ東のモデル化図

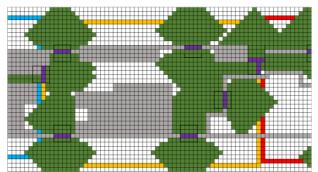




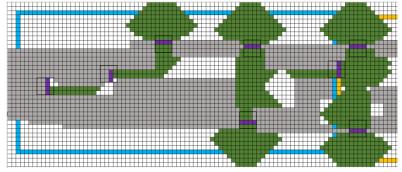
- (c) シャミネ西のモデル化図
- 図. JR 松江駅のモデル化図



(a) シャミネ東のセンサ有効範囲図



(b) コンコースのセンサ有効範囲図



- (c) シャミネ西のセンサ有効範囲図
- 図. JR 松江駅のセンサ有効範囲図

4.2.2.対象者設定

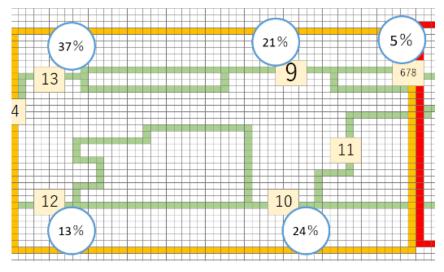
サービス時間の人流センサデータ数の一日当たりの平均を表に示す.「センサ通過のべ人数」とは in/out データ両方を加算した値である.

センサ番号	1	2	3	4	5			
センサ通過のべ人数	0	0	2	0	2			
センサ番号	6	7	8	9	10			
センサ通過のべ人数	94	14	68	452	514			
センサ番号	11	12	13	14	15			
センサ通過のべ人数	1181	277	811	0	×			
センサ番号	16	17	18	19	総和			
センサ通過のべ人数	0	0	0	0	3415			

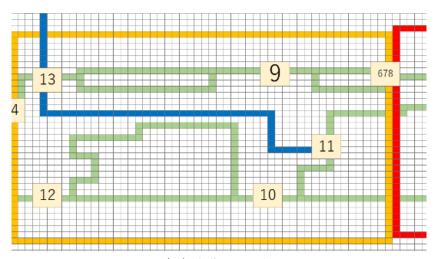
表:登校時人流センサ毎の総データ量



ある人物が駅を通過する際には少なくとも駅への流入時と流出時の 2 つのデータを生成すると考えられる.この仮定と表から、一日当たりの 7 時から 8 時の駅利用者をおおよそ1700 人分とする.鉄道利用者に占める学生の割合は全体の 7%程度[9]であるから、対象者の最大人数を全体の10%程度とし、200 人と設定した.対象者が駅を通過する際に使う流入口は人流センサ 11 番(改札)で固定とし、流出口は物理的に駅の外側に接している人流センサから表に従い確率で選択した.利用される流出口の利用率を図に示す、人流センサ 11 番から流入し、人流センサ 13 番から流出する際の対象者のルートを図に青色で示す.



流出口使用割合



対象者移動ルート例

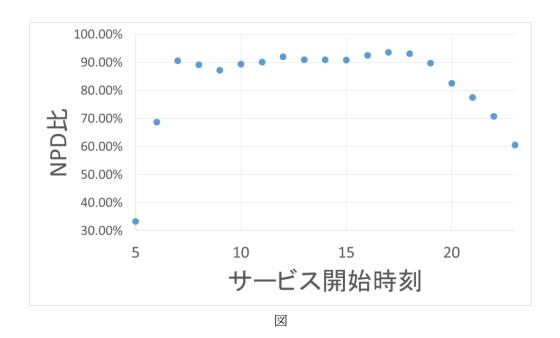
4.3. シミュレーション内容

シミュレーションを行い、クラウドサービスによって提供される松江駅登校者見守りサービスにおいて、サービス開始時刻毎のNPD流出比についてのグラフ、センサ毎のNPD流出比についてのグラフ、そして対象者人数毎のNPD流出比のグラフを導出した.

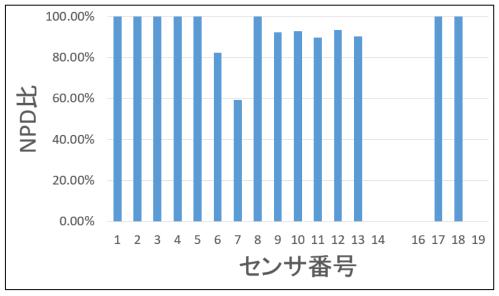
LPLC モデルによって提供される松江駅登校者見守りサービスにおいて、ルータアクセス 可能距離の値を設定し、対象者数を1人に加え、25人から200人まで25人刻みで変化させ シミュレーションを行った.また、対象者を1人とした時にルータアクセス可能距離を変化 させた場合のNPD 流出比をシミュレーションした.

4.4. シミュレーション結果

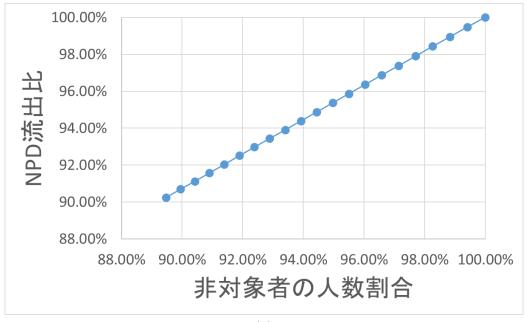
クラウドサービスによって 5 時から 24 時の間, 1 時間毎に 200 人の見守りサービスを行った際の NPD 流出比を算出した. サービス開始時刻毎の NPD 流出比のグラフを次に示す.



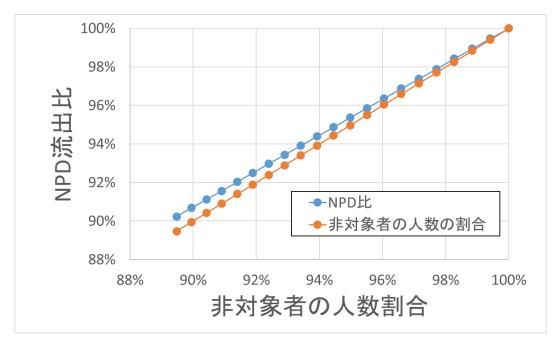
クラウドサービスによって登校者見守りサービスをシミュレーションした際のセンサ毎の NPD 流出比グラフを次に示す. 3 章 1 節で触れた通り, 人流センサ 15 番は無視している.



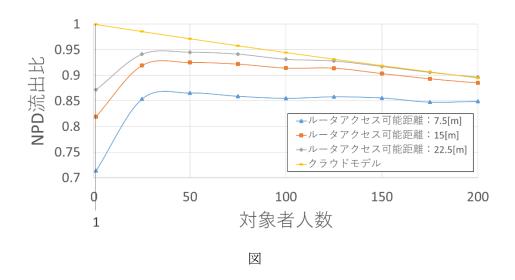
クラウドサービスによって登校者見守りサービスにおいて対象者を0人から200人まで、10人刻みで変動させた際のNPD流出比のグラフを次に示す.



义



ルータアクセス可能距離毎の対象者人数と NPD 流出比の関係, およびクラウドモデルの NPD 流出比を図にまとめた.



表

ルータアクセス可能距離	対象者人数が1人の際の NPD 流出比
2	
4	
6	
8	
10	
12	
14	
16	
18	
20	
22	
24	

NPD 流出比が小さいほどプライバシ保護能力が高いといえるので、ルータアクセス可能 距離に依らず、クラウドモデルよりも LPLC モデルの方が優れている.

4.5. 考察

7時から20時の間絶えずプライバシデータ流出比が80%を下回らないことがわかる.早朝,深夜に学生見守りサービスを行うのであればNPD流出比は比較的小さくなるが,その時間に学生が駅を利用するとは考えにくい.

センサ毎に NPD 流出比に差があることが確認できる. 特に NPD 流出比が 100%である 1, 2, 3, 4, 5, 8, 17, 18番の人流センサはこのサービスでは不要になる. もしこのセンサをサービスから除くことができれば、NPD 流出比は 90.50%から 90.29%と僅かながら確実に減少する. その様子を表に示す.

表

センサ番号	NPD 量	対象者データ量	NPD 流出比
1	52	0	100.00%
2	52	0	100.00%
3	420	0	100.00%
4	43	0	100.00%
5	491	0	100.00%
6	23832	5146	82.24%
7	3732	2573	59.19%
8	17328	0	100.00%
9	93567	7921	92.20%
10	101864	7992	92.73%
11	297792	34599	89.59%
12	61536	4388	93.34%
13	180524	19416	90.29%
14	0	0	0.00%
15	-	-	- %
16	0	0	0.00%
17	51	0	100.00%
18	48	0	100.00%
19	0	0	0.00%
全センサの合計	781332	82035	90.50%
100%以外	7,699,47	0000	00.900/
センサ合計	762847	82035	90.29%
削減したデータ	18485	0	0.21%

200名の学生を見守りサービスの対象に設定しても NPD 流出比は 90%より大きいことがわかる. また NPD 流出比が対象者人数に対して線形に減少していることから,駅利用者全体の人数と第三者人数の比がそのまま NPD 流出比を表していると予想できる. 4.2.2 で述べた通り対象者 200 人は駅利用者のおよそ 10%であったが,その時の NPD 流出比はおよそ90%である.

ルータアクセス可能距離がセンサ有効距離より短いと追跡サービスとして不十分であるが、反対にルータアクセス可能距離があまりに長いと NPD センシング過多の原因となる。よって、センサ有効範囲を内包しつつルータアクセス可能距離を短くすることにより、プライバシに配慮した追跡サービスが実現できる。また図から、対象者数が少ないときの LPLC モデルのプライバシ保護能力が特に優れているとわかる。そのため対象者人数に対してLPLC ノードの数を減らすことが出来れば、より NPD 流出比を抑制できる。

実際のサービス領域は対象者の映りにくい領域までカバーする必要がある。しかし対象者が映っていないセンサは NPD 収集を過度に行う原因にもなる。そのようなセンサを排除すると NPD 量と NPD 流出比の両方を小さくできると上述したが、センサデータを収集してから対象者と判断するクラウドモデルでは上記の方式は実装困難である。

センシング可能範囲と NPD 収集量の間にはトレードオフの関係がある. ただ NPD 比の削減のみを目指し続けると, そもそもセンシングをしなければよいという支離滅裂な回答が得られかねない. そこでサービス領域カバー率という指標を用意する. これはサービス領域の面積の内何%がセンシング可能であるかを示す指標である. これが 100%に近いほど, 見守りサービスとして十分であると言える. 本シミュレータは図と図より, サービス領域カバー率は 46.1%であった.

第5章 まとめ

序章であげた社会問題に対応する.

本稿ではクラウドモデル及び LPLC モデルの NPD 流出比を数理モデルで表現し、シミュレータからクラウドモデルと LPLC モデルを比較検討した。その結果、LPLC モデルはクラウドモデルよりもプライバシ保護能力に優れているとわかった。

今後はより多人数を見守る際のNPD流出比の削減方法,及びLPLCモデルの実証実験等を行い、LPLCモデルの実用化に向け研究を進めていく.

謝辞

本稿の執筆にあたり、熱心にご助言を下さった指導教員の干川 尚人教授に感謝を申し上げます. 千葉大の方々. また、数理モデル作成にあたり渡邉 扇之介教授には多大なアドバイスとアイデアを頂きました. 感謝を申し上げます. オープンデータについての質問に親身にご回答いただきました G 空間情報センターの角田さんにも感謝を申し上げます.

参考文献

- [1] 総務省. 特集 データ主導経済と社会変革. 平成 29 年版 情報通信白書 ICT 白書 2017, 第1部, 第2章, pp. 79.
- [2] 大橋 正良ら, "ユビキタス環境におけるプライバシ保護の一検討",電子情報通信学会 総合大会 B-20-53 2011.
- [3] 干川尚人,下馬場朋禄,伊藤智義,"地産地消型アーキテクチャによるセンサネットワークデータのプライバシ保護,"情報処理学会論文誌, vol.59, No.12, pp.2180-2190, Dec, 2018.
- [4] "AWS IoT プラットフォーム". https://aws.amazon.com/jp/s/dm/optimization/sem/iot-signup/ (参照 2018-12-20)
- [5] "PaaS で誰もがデータ分析". 日経コンピュータ. http://tech.nikkeibp.co.jp/it/atclact/active/15/010700161/010700003/(参照 2018-12-20)
- [6] 干川尚人,下馬場朋禄,伊藤智義,"地産地消型アーキテクチャによるセンサネットワークデータのプライバシ保護,"情報処理学会論文誌, vol.59, No.12, pp.2180-2190, Dec, 2018.
- [7] 社会基盤情報流通推進協議会、人流解析チーム、"松江駅構内人流センサデータ"
- [8] "株式会社アルコム RD-CA213 AHD カメラ 220 万画素バレット型赤外線搭載屋外カメラ" https://www.arucom.ne.jp/store/g/g250-011409/ (参照 2018-12-20)
- [9] "株式会社アルコム 人相確認距離から防犯カメラを選ぶ。" https://www.arucom.ne.jp/sample/ (参照 2018-12-20)
- [10] 関東交通広告協議会, "交通広告調査レポート 2009," 鉄道利用者プロフィール, pp39